

BAB 3

METODE PENELITIAN

3.1 Objek penelitian

Penelitian tugas akhir ini dilakukan di Universitas Dian Nuswantoro Semarang Jl. Nakula I No. 5 – 11, Pendrikan Kidul, Semarang Tengah, Semarang.

3.2 Metode Pengumpulan Data

Dari penelitian yang dilakukan, data sekunder tidak termasuk dalam penelitian ini karena data yang di ambil bukan dari data publik melainkan data primer karena data yang diperoleh bersifat privat atau pribadi. Sumber data yang akan digunakan adalah dataset mahasiswa Fakultas Ilmu Komputer Universitas Dian Nuswantoro tahun 2008 sampai 2012 yang telah dinyatakan lulus dan diwisuda. Data yang di peroleh melalui *softcopy* dan dikirim melalu email mahasiwa. Jumlah data yang di peroleh sebelum proses *cleaning* dapat di lihat pada tabel berikut :

Tabel 3.1 Data set sebelum proses *cleaning*

No	Program Studi	Jumlah Record
1.	Teknik Informatika – S1	1807
2.	Sistem Informasi – S1	1060
3.	Desain Komunikasi Visual	459
4.	Teknik Informatika – D3	346
5.	Broadcasting – D3	204

Berikut ini adalah data sampel mahasiswa :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	id_mhs	jurusan	angkatan	wisuda_ke	jurusan	rata_uan	gel_daf	nm_sita	nama_prov	nama_kota	ipk	rata_nilai_1	rata_nilai_2	rata_nilai_3	rata_nilai_4	rata_nilai_5	rata_nilai_6
2	2125	A21	2010	45	W	W	W		W	W	2.67	W	W	W	W	W	W
3	2201	A21	2008	42	W	0.00	W	SMU KESATRIAN 2	W	W	2.82	W	W	W	W	W	W
4	3292	A24	2008	46	W	0.00	W	SMA NEGERI 04 JAYAPURA	PAPUA	KOTA JAYAPURA	2.83	79.31	82.33	77.07	76.84	0.00	0.00
5	3908	A12	2008	49	W	0.00	W	SMA 5 SEMARANG	JAWA TENGAH	KOTA SEMARANG	3.05	66.44	69.10	70.31	68.85	0.00	0.00
6	4773	A11	2010	46	W	W	W	SMA 3 Semarang	W	W	3.16	66.37	61.25	64.00	66.35	0.00	0.00
7	5408	A11	2010	49	W	W	W	SMUN 06 Semarang Barat	W	W	2.8	76.12	79.12	78.71	78.85	0.00	0.00
8	6515	A12	2012	56	IPA	0.00	PMDKII	1865	JAWA TENGAH	KOTA SEMARANG	3.69	73.29	76.29	78.38	79.84	0.00	0.00
9	9338	A11	2011	53	W	0.00	KHUSUSI	121	JAWA TENGAH	KABUPATEN BOYOLALI	3.71	74.52	76.00	77.69	79.69	0.00	W
10	9566	A14	2008	50	W	0.00	W	MA An-najah Bekasi	JAWA TENGAH	KABUPATEN BREBES	3.19	83.76	83.12	78.57	80.14	0.00	0.00
11	9567	A14	2008	53	W	0.00	W	SMK NEGERI 2 SLAWI	JAWA TENGAH	KABUPATEN TEGAL	3.04	65.35	66.80	70.60	72.00	0.00	0.00
12	10794	A24	2008	42	W	0.00	W	SMA Theresiana	W	W	3.05	64.70	67.33	66.50	72.00	0.00	0.00
13	10796	A11	2008	50	W	0.00	W		JAWA TENGAH	KOTA SEMARANG	2.68	68.00	69.00	72.00	72.90	0.00	0.00
14	10798	A11	2008	45	W	0.00	W	SMA MAGEH 2 PSAN	W	W	3.43	80.33	81.31	82.00	83.07	84.69	W
15	10799	A11	2008	45	W	0.00	W	SMA N 14	W	W	2.88	70.75	71.35	75.46	76.07	0.00	W
16	10800	A11	2008	51	W	0.00	W		JAWA TENGAH	KABUPATEN SEMARANG	2.66	79.65	76.33	79.09	82.35	0.00	0.00
17	10801	A11	2008	45	W	0.00	W	SMA SEMESTA	W	W	3.45	77.16	75.55	74.82	73.61	72.55	W
18	10803	A11	2008	50	W	0.00	W	SMA Theresiana 1	JAWA TENGAH	KOTA SEMARANG	2.69	78.62	80.25	77.76	77.53	0.00	W
19	10804	A11	2008	46	W	0.00	W	SMA KEBON DALAM	W	W	3.04	71.96	72.33	74.11	74.86	0.00	W
20	10805	A11	2008	45	W	0.00	W	Sedes Sapientiae	W	W	2.98	71.70	72.40	68.70	72.80	69.50	W
21	10806	A11	2008	45	W	0.00	W	SMA N 1 Kendal	W	W	3.14	75.93	80.13	81.42	81.25	0.00	0.00
22	10807	A11	2008	45	W	0.00	W		W	W	2.95	69.39	71.12	71.76	73.33	0.00	W
23	10808	A11	2008	45	W	0.00	W	SMA Kristen YSKI	W	W	3.6	68.90	68.96	71.38	72.95	0.00	W
24	10809	A11	2008	45	W	0.00	W	KRISTA MITRA	W	W	3.37	63.81	60.37	61.00	64.30	0.00	W
25	10810	A11	2008	45	W	0.00	W	SMA PL Don Bosko	W	W	3.5	75.44	77.79	75.85	77.93	0.00	0.00
26	10811	A11	2008	47	W	0.00	W		JAWA TENGAH	KABUPATEN KUDUS	3.04	61.05	69.18	64.64	72.00	70.84	0.00

Gambar 3.1 Sampel Data Mahasiswa

Dari gambar 3.1 terlihat bahwa data Mahasiswa Fakultas Ilmu Komputer angkatan 2008 – 2012 terdiri dari 17 atribut.

Penjelasan nama variabel dari data tersebut adalah :

Tabel 3.2 Atribut Data Awal

No	Nama Atribut	Keterangan
1	id_mhs	Nomor Induk Mahasiswa
2	program_studi	Program studi mahasiswa
3	Angkatan	Tahun masuk mahasiswa
4	wisuda_ke	Periode wisuda mahasiswa
5	Jurusan	Jurusan pada sekolah menengah mahasiswa
6	rata_uan	Rata-rata nilai UN mahasiswa

No	Nama Atribut	Keterangan
7	gel_daf	Gelombang daftar mahasiswa
8	nm_slta	Nama SLTA mahasiswa
9	nama_prov	Nama provinsi mahasiswa berasal
10	nama_kota	Nama kota mahasiswa berasal
11	Ipk	Indeks Prestasi mahasiswa pada saat wisuda
12	rata_nilai_1_1	Rata-rata nilai mahasiswa pada SLTA kelas X semester 1
13	rata_nilai_1_2	Rata-rata nilai mahasiswa pada SLTA kelas X semester 2
14	rata_nilai_2_1	Rata-rata nilai mahasiswa pada SLTA kelas XI semester 1
15	rata_nilai_2_2	Rata-rata nilai mahasiswa SLTA kelas XI semester 2
16	rata_nilai_3_1	Rata-rata nilai mahasiswa pada SLTA kelas XII semester 1
17	rata_nilai_3_2	Rata-rata nilai mahasiswa pada SLTA kelas XII semester 2

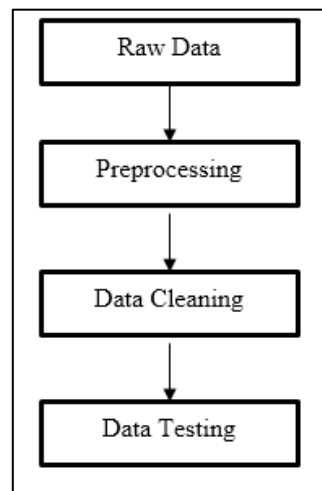
3.3 Implementasi Algoritma Apriori

Dari kajian pustaka atau *study literature* penelitian pada bab 2, Algoritma Apriori berpotensi untuk menyelesaikan masalah. Berikut adalah langkah penerapan algoritma apriori :

3.3.1 Preprocessing

Sebelum menerapkan Algoritma Apriori , tahap *preprocessing* terlebih dahulu dilakukan . Data yang tidak berkualitas akan menghasilkan *data mining* yang tidak berkualitas. Tidak semua data dan atribut dari data mahasiswa akan digunakan untuk diolah, sebab penilitian yang akan dilakukan memiliki batasan – batasan data. Guna mendapatkan data yang berkualitas diperlukan tahap *preproccesing*.

Berikut alur tahapan *preprocessing*:



Gambar 3.2 Tahapan Preprocessing

1. *Data Cleaning*

Data Cleaning mempunyai beberapa tugas, antara lain :

- a. Mengisi beberapa nilai yang hilang, data yang tidak lengkap (*missing value*)
- b. Menghilangkan *outliers* serta memperhalus *data noise*.
- c. Data yang tidak konsisten harus diperbaiki.
- d. Memecahkan redudansi yang disebabkan oleh integrasi data.

Berikut adalah jumlah data yang merupakan data *testing* atau data yang sudah siap diolah dengan menggunakan algoritma *apriori* :

Tabel 3.3 Data set yang siap diolah

No	Program Studi	Jumlah Record
1.	Teknik Informatika – S1	352
2.	Sistem Informasi – S1	228
3.	Desain Komunikasi Visual	125
4.	Teknik Informatika – D3	68
5.	Broadcasting – D3	47

Beberapa atribut akan dihapus karena tidak mempengaruhi dalam penelitian ini.

Gambar 3.3 Penghapusan atribut yang tidak digunakan

Gambar 3.3 diatas adalah penghapusan atribut – atribut dari data set utama yang tidak digunakan dalam proses penelitian ini karena tidak berkaitan dengan proses perhitungan algoritma *apriori* yang akan menghasilkan satu set acuan yang berisi informasi bagi calon mahasiswa untuk mengambil keputusan dalam mengambil program studi.

	A	B	C	D	E	F	G	H	I
1	id_mhs	program_studi	angkatan	wisuda_ke	jurusan	gel_daf	ipk	rata_nilai_1_1	rata_nilai_1_2
2	2125	A21	2010	45	\N	\N	2.67	\N	\N
3	2201	A21	2008	42	\N	\N	2.82	\N	\N
4	3292	A24	2008	46	\N	\N	2.83	79.31	82.33
5	3908	A12	2008	49	\N	\N	3.05	66.44	69.20
6	4773	A11	2010	46	\N	\N	3.26	66.37	61.25
7	5408	A11	2010	49	\N	\N	2.8	76.12	79.12
8	6515	A12	2012	56	IPA	PMDKII	3.69	73.29	76.29
9	9338	A11	2011	53	\N	KHUSUSI	3.71	74.52	76.00
10	9366	A14	2008	50	\N	\N	3.19	83.76	83.12
11	9567	A14	2008	53	\N	\N	3.04	65.35	66.80
12	10794	A24	2008	42	\N	\N	3.05	64.70	67.33
13	10796	A11	2008	50	\N	\N	2.68	68.00	69.00
14	10798	A11	2008	45	\N	\N	3.43	80.33	81.31
15	10799	A11	2008	45	\N	\N	2.88	70.75	71.35
16	10800	A11	2008	51	\N	\N	2.66	79.65	76.33
17	10801	A11	2008	45	\N	\N	3.45	77.16	75.55
18	10803	A11	2008	50	\N	\N	2.69	78.62	80.25
19	10804	A11	2008	46	\N	\N	3.04	71.96	72.33
20	10805	A11	2008	45	\N	\N	2.98	71.70	72.40
21	10806	A11	2008	45	\N	\N	3.14	75.93	80.13

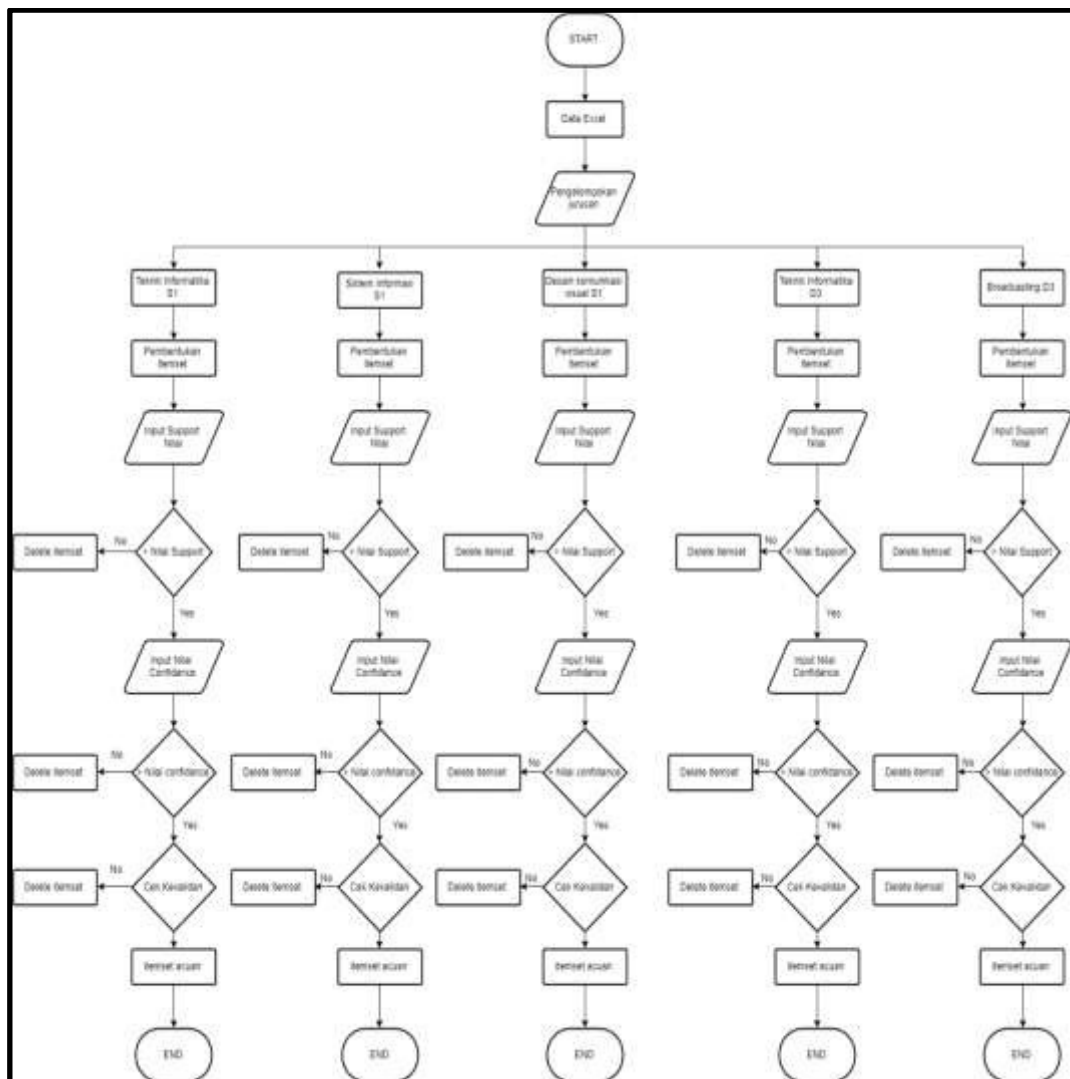
Gambar 3.4 Hasil dari proses data selection

Gambar 3.4 menunjukkan hasil dari *attribute selection* yang sudah disesuaikan dengan kebutuhan proses penelitian. Definisi dari *attribute selection* adalah dihapusnya atribut – atribut yang tidak diperlukan untuk proses penelitian agar dapat memudahkan dalam mengolah data.

Tabel 3.4 Atribut data testing

No	Nama Atribut	Keterangan
1	id_mhs	Nomor Induk Mahasiswa (sebagai primary key)
2	program_studi	Program studi mahasiswa
3	Angkatan	Tahun masuk mahasiswa
4	wisuda_ke	Periode wisuda mahasiswa
5	Jurusan	Jurusan pada sekolah menengah mahasiswa
6	gel_daf	Gelombang daftar mahasiswa
7	Ipk	Indeks Prestasi mahasiswa pada saat wisuda
8	rata_nilai_1_1	Rata-rata nilai mahasiswa pada SLTA kelas X semester 1
9	rata_nilai_1_2	Rata-rata nilai mahasiswa pada SLTA kelas X semester 2

3.3.2 Penerapan Algoritma Apriori



Gambar 3.5 Proses Penerapan Algoritma Apriori dalam Penelitian

Analisis asosiasi merupakan dasar dari teknik data mining lainnya . Analisis pola frekuensi tinggi (*frequent pattern mining*) adalah salah satu tahap Analisis asosiasi yang banyak digunakan oleh para peneliti untuk menghasilkan beberapa algoritma yang efisien. 2 tahap analisis asosiasi adalah :

1. Analisis pola nilai penunjang (*support*)

Dalam mencari kombinasi item yang memenuhi syarat minimum dari nilai pendukung yang ada di dalam database dilakukan pada tahap ini.

Rumus nilai support :

$$\text{Support}(A) = \frac{\text{jumlahTransaksiMengandung}A}{\text{NilaiTransaksi}} \quad (3.1)$$

Rumus nilai support 2 item :

$$\begin{aligned} \text{Support}(A, B) &= \text{Support}(A \cap B) = \\ &\frac{\text{jumlahTransaksiMengandung}A \text{ dan } B}{\text{NilaiTransaksi}} \end{aligned} \quad (3.2)$$

2. Perhitungan nilai kepastian (*confidence*)

Aturan asosiasi ditentukan setelah pola frekuensi tinggi ditemukan . Aturan asosiasi yang dicari adalah aturan assosiatif yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiasi “ jika A maka B “. Rumus mencari Nilai confidence dari aturan “ jika A maka B “ :

$$\begin{aligned} \text{confidence} &= \text{Support}(A|B) = \\ &\frac{\text{jumlahTransaksiMengandung}A \text{ dan } B}{\text{jumlahTransaksi mengandung}A} \end{aligned} \quad (3.3)$$

3.3.2.1 Langkah – Langkah Proses Aturan Asosiasi

Berikut adalah langkah – langkah dalam proses aturan asosiasi:

1. Mencari 1-itemset (himpunan item yang terdiri dari 1 item) dan menghitung nilai supportnya dengan cara men-scan database. Lalu bandingkan nilai support yang di dapatkan dengan nilai minimum support yang ditentukan terlebih dahulu . a itemset tersebut termasuk dalam large itemset jika nilainya lebih besar atau sama dengan minimum support.
2. Hanya Itemset termasuk dalam large itemset yang diikutkan dalam iterasi selanjutnya.
3. Hasil large itemset pada iterasi pertama (L1) digunakan pada iterasi kedua sistem untuk membentuk kandidat itemset kedua (L2). Hasil large itemset pada iterasi selanjutnya digunakan untuk iterasi selanjutnya menggunakan hasil large itemset pada iterasi sebelumnya (Lk-1) untuk membentuk kandidat itemset berikut (Lk). Sistem akan menggabungkan (join) Lk-1 dengan Lk- untuk mendapatkan Lk, seperti pada iterasi sebelumnya hanya Itemset termasuk dalam large itemset yang diikutkan dalam iterasi selanjutnya.
4. Itemset hasil join kemudian dihitung supportnya
5. Proses join akan membentuk kandidat yang akan terus dilakukan hingga himpunan kandidat itemsetnya null, atau sampai kandidat sudah tidak ada yang bisa dibentuk lagi.
6. Kemudian, association rule yang memenuhi nilai support dan confidence yang telah ditentukan dibentuk dari frequent itemset tersebut
7. Nilai yang memiliki kemiripan di anggap sebagai satu data dalam association rule
8. Nilai minimum harus terpenuhi dalam pembentukan association rule
9. Cari himpunan bagian L yang tidak kosong untuk setiap large itemset L. Dihasilkan rule dengan bentuk $aB(L-a)$ jika supportnya (L) dan supportnya (a) lebih besar dari minimum support Untuk setiap himpunan bagian tersebut.

3.4 Evaluasi (*Evaluation*)

Lift ratio adalah parameter penting setelah mendapatkan nilai support dan confidence dalam association rules yang digunakan untuk mengevaluasi sebuah aturan asosiasi. *Lift ratio* adalah nilai yang menunjukkan kevalidan proses transaksi dan memberikan informasi apakah benar produk A dibeli bersamaan dengan produk B. *Lift Ratio* dapat dihitung dengan rumus sebagai berikut :

$$Lift\ ratio = \frac{Support\ (A \cap B)}{Support\ (A) * Support\ (B)} \quad (3.4)$$

Support (A \cap B) = nilai support yang mengandung transaksi A dan transaksi B

Support (A) = nilai support yang mengandung transaksi A

Support (B) = nilai support yang mengandung transaksi B

Jika nilai lift ratio > 1 maka artinya ada manfaat dari aturan tersebut. Semakin tinggi nilai lift ratio maka semakin besar kekuatan asosiasinya. Yang berarti bahwa produk A benar-benar dibeli bersamaan dengan produk B. [13]